View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Health Risk of Heating Fuel Choice: A Simultaneity Causality Analysis

Zheng Liu
Ph.D., Research Assistant
Department of Agricultural Economics/ University of Kentucky
E-mail: liuzkm@uky.edu

Angelos Pagoulatos
Professor
Department of Agricultural Economics/ University of Kentucky
E-mail: apagoula@uky.edu

Wuyang Hu
Assistant Professor
Department of Agricultural Economics/ University of Kentucky
E-mail: wuyang.hu@uky.edu

Selected Paper prepared for presentation at the Southern Agricultural Economics Association Annual Meeting, Orlando, FL, February 6-9, 2010

Copyright 2010 by Zheng Liu, Angelos Pagoulatos and Wuyang Hu. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Health Risk of Heating Fuel Choice: A Simultaneity Causality Analysis

Abstract

Combustion-generated pollutants, principally those from solid-fuels including biomass and coal when cooking and heating, bring out a significant public health hazard in both developed and developing countries. Most of the existing studies addressing this issue focus on developing countries, and on exposure when cooking rather than heating. By using Kentucky rural data, this research explores the health risk associated with heating fuel choice. Given the simultaneity between heating fuel choice and prevalence of asthma and allergy, we obtain the instrumental variable (IV) estimate for Logit models through the Generalized Method of Moments (GMM). After correcting for simultaneity bias, we do not find strong evidence supporting the causal relationship between polluting heating use and the prevalence of asthma, allergy, and other respiratory disease. Some demographic and lifestyle factors do have significant effects on the prevalence of these diseases.

Keywords: *combustion-generated pollutants, indoor air pollution, heating fuel choice, health risk, GMM-IV Estimation*

Introduction

Indoor air pollution (IAP) is a public health problem in both developed and developing countries (Ezzati et al., 2003).Since late of 1980's, based on comparative risk studies, EPA and its Science Advisory Board (SAB) have consistently ranked indoor air pollution among the top five environmental risks to public health (EPA, 1993).

Among the four components of indoor pollution (combustion products, chemicals, radon, and biologic agents), combustion-generated pollutants, principally those from solid-fuel such as biomass (wood, dung and crop residues) and coal used in cooking and heating, have been the focus of epidemiologic and physiologic research, especially in developing countries. Biomass or coal smoke contains a large number of pollutants that have known health hazards: particulate matter (PM), carbon monoxide (CO), nitrogen dioxide, sulfur oxides (mainly from coal), formaldehyde, and polycyclic organic matter, including carcinogens such as benzopyrene and benzene (Ezzati and Kammen, 2002b). Based on the reviews by Smith et al (2000) and Bruce et al (2000), the focus of the epidemiological research addressing the health hazards of Indoor Air Pollution (IAP) from solid fuel combustion is given to acute (lower) respiratory infections (ALRI), chronic obstructive pulmonary disease (COPD), and lung cancer (due to coal) for which the evidence is the most robust.

Although developing countries are typical research locations, the health concern of solid fuels combustion still exists in developed countries. People experienced coal and wood for heating and cooking at their younger age are now in the risk age range for respiratory diseases and lung cancer, especially those who grew up in rural areas. Although dirty fuels are no longer dominant, in the past decade, there has been some increased consumption of wood for heating and cooking in developed countries, to create mood-setting atmosphere or grill food (Zerbe, 2004). In the U.S., from the RECS data (2005), about 2.9 million households (2.6% of the total housing units) use wood as the main heating fuel and about 79% of them live in rural areas. About 8.9 million households (8% of the total housing units) use wood as their secondary heating fuel.

Although the U.S. Environmental Protection Agency (EPA) has promulgated New Source Performance Standards (NSPS), which set up particulate emission standard for wood heaters to be certified. However, by 1998, only about 11% of the wood stoves in use were EPA certified, and only 4% of the fireplace inserts were EPA certified (Houck et al., 1998).The EPA and the U.S. Consumer Product Safety Commission (2007) also claimed that use of unvented combustion appliances (such as kerosene and oil fueled space heaters) in closed settings may also be associated with health risks because of exposure to polluting emissions.

According to the review report by the U.S. Energy Information Administration, residential coal consumption increased by 9 percent in 2007 and keeps the high level in 2008 (Freme, 2009). It is necessary to explore the health risk associated with the use of wood, coal burning and other unvented combustion appliance that are fueled by kerosene or fuel oil for heating. In particular, we are interested on whether over time awareness due to information availability in the U.S. regarding the adverse health effects of different heating fuels may have prompted a strong averting activity. Averting activity represents the activities undertaken to avoid or reduce exposure to pollution. It includes change of heating choice, such as shifting away from using polluting heating fuel or correctly using combustion appliances which are certified by the EPA and follow the safety guidelines strictly. The potential effect of averting activity may not allow the observation of the health risk associated with heating fuel choice in U.S. when exploring cross-sectional data.

Most of the existing studies focus on developing countries, and on exposure when cooking rather than heating. Using the Kentucky Homeplace Program survey data, this paper explore the health risk of heating fuel choice in rural Kentucky. To find the relationship between heating fuel choice and the occurrence of disease from a panel set of observations, it would suggest that people do not engaged in averting activity. If we can not exclude the averting behavior, especially when we use cross-sectional data to conduct the analysis, we will consider the simultaneity bias produced by the averting activity and obtain the unbiased estimate for the effect of heating choice on the occurrence of disease.

Model

Choice model will be used to identify the exposure–response relationship between the use of polluting heating fuel and the prevalence of some disease. This exposure–response relationship exists conditional on the effects of individual's behavior (averting activity). The basic equation we will consider is:

Prob ($disease$=1) = f (ph, $X \setminus AVERT$)

Where *disease* is a dummy variable, which is equal to 1 if individual has some specific disease, equal to 0 otherwise. *ph* is also a dummy variable and equal to 1 if individual use some polluting heating fuel, equal to 0 otherwise. *X* is a set of other exogenous variables (such as social-economic and demographic, lifestyle and other exogenous explanatory variables) that maybe influence the occurrence of some specific disease. *AVERT* indicates the existence of averting activity. *AVERT* could be ≥ 0. If *AVERT >0,* we should consider the simultaneity problem associated and correct the simultaneity bias when estimating the exposure-response relationship between the use of polluting heating fuel and the prevalence of the disease.

To address this simultaneity problem, we use the instrumental variable estimation (IVE). In the linear model, the most common form of IVE is two-stage least squares (2SLS). However for the non-linear discrete model in our case, standard 2SLS procedure is not readily available and the derivation of the IV estimator is not trivial. The literature on this issue is sparse. In this paper, we will obtain the IVE for simultaneous logistic regression by using the Generalized Method of Moment (GMM).

GMM Estimation of a non-linear model (such as Logit Model) is based on the similar intuition in a linear model (Hayashi, 2000; Davidson and MacKinnon, 1993; Foster, 1997). Consider a classic Logistic regression, where

$$\Pr(y_t = 1) = \frac{e^{X_t \beta}}{(1 + e^{X_t \beta})} \quad \text{or} \quad (1 + e^{-X_t \beta})^{-1} \tag{1}$$

y_t is a binary variable, which equals to1 if an individual *t* has some event and 0, otherwise. *X* is a N × K matrix of regressors (*K* is the total number of explanatory variables). X_t is a 1×K vector and represents the *t*th row of the *X* matrix. Error term is defined as $(y_t - (1 + e^{-X_t \hat{\beta}})^{-1}$ Based on the orthogonality conditions between the explanatory variable and the error term, the estimator-defining equations can be set as follows.

$$\frac{1}{N}\sum_{t=1}^{N} x_{t,k}(y_t - (1 + e^{-X_t \hat{\beta}})^{-1} = 0 \tag{2}$$

Where k =1 to K. X_t is defined as above. $x_{t,k}$ is the value of the *k*th explanatory variable for the *t*th individual. GMM estimates $\hat{\beta}$ can be obtained by solving these equations as long as the number of moment conditions is equal to the number of parameters to be estimated. Maximum-likelihood estimation (MLE) is a special case of GMM in this situation because the moments are the first-order conditions for maximizing the log-likelihood function (Greene, 1993).

IV estimates in the non-linear model can be obtained in the same manner as they are in the linear case: by replacing the endogenous regressors in the estimator-defining equations with appropriate instruments (Amemiya, 1985). In the case of logistic regression, this produces the following equation:

$$\frac{1}{N}\sum_{t=1}^{N} w_{t,k}(y_t - (1 + e^{-X_t \tilde{\beta}})^{-1} = 0 \tag{3}$$

where k= 1 to *K*. $w_{t,k}$ represents the value of the *k*th instrument for the *t*th individual. Equation (3) is also implied by the first-order conditions for instrumental variable estimation. The corresponding estimating equations of (3) and (2) are identical for the regressors which are not endogenous variable (they can serve as their own instruments).

GMM estimates $\tilde{\beta}$ can be obtained by solving the set of *K* equations when the model is just identified (there is one instrument per variable). In the over-identified model, GMM estimate can be obtained by minimizing criterion function like equation (4)

$$\varepsilon^T W (W^T W)^{-1} W^T \varepsilon \tag{4}$$

where the residual ε is defined as $(y_t - (1 + e^{-X_t \tilde{\beta}})^{-1}$. If *L* is the number of instruments ($L > K$), ε is the $N\times 1$ vector of residuals, *W* is the $N\times L$ matrix of instruments. GMM estimates for just-identified model can also be obtained by this approach and the minimized value of the criterion function is 0.

In a dichotomous outcome model, as in this case, the residuals are expected to be heteroskedastic. To improve the efficiency of the estimates, we can modify the criterion function (4) and minimize criterion function like equation (5)

$$\varepsilon^T W (W^T \Omega W)^{-1} W^T \varepsilon \tag{5}$$

where Ω is the variance–covariance matrix of the error term. $(W^T \Omega W)$ need to be estimated before minimizing criterion function (5). White's (1980) heteroskedasticity-consistent covariance matrix estimator can accommodate flexible forms of heterogeneity in $(W^T \Omega W)$. By this way, an estimate of the weighting matrix is obtained in the first stage. It plugged into the objective function (5) and function (5) is minimized to find the GMM estimates.

Data

The data used in this study come from the health survey data of the Kentucky Homeplace Program (KHP). The surveys are the initial interview with respondents who want to enroll in the Kentucky Homeplace Program. The purpose of this program is to help patients who live in rural areas to find and use the services they are qualified, provide preventive services for some chronic diseases and collect data for long-term studies of illness prevalence in the area.

The dataset used in this study are KHP health survey data (cross-sectional data) collected in 2005 and 2006. Besides the demographic, social-economic and risk factors information, the key question for this study is: “What type of heat do you have?” Respondents may choose more than one type of heating fuel from electric, gas, coal, wood, fuel oil, kerosene, and others (then give any comments using the space given in the survey). In descending order, the percentage of heating fuel used by the sample housing units are electricity (66.8%), gas (29.9%), wood (7%), kerosene (3.8%),coal (3.4%), fuel oil (0.6 %), and other fuel (0.2%).

Table 1 reports the definitions and statistics of the variables used in the study. The average age of respondents is 53 years old, which is higher than reported median age of 35.9 for Kentucky residents (Kentucky Demographics 2005). 37% of the respondents are male and 95% of the sample are white (not Hispanic or Latino). The average length of education is 11 years. The average annual income is about $12,717 which is much lower than the state average of $40,299 a year and national average of $50740 (U.S Census

Bureau, 2007). These statistics are relevant to the geographic service area and the service objective of the Kentucky Homeplace Program. The program focuses on the rural counties of Kentucky and most of the clients (respondents of the survey) are retirees and lower income receivers. About 45 % of the respondents participate in physical activities, and 53% have used tobacco products.

For the heating fuel choice from the survey, two categories were created and used in this study: Non-Polluting Heating (*nph*) and Polluting Heating (*ph*). The former includes heating fuel choices of electric and gas. The latter includes heating choices of coal, wood, fuel oil, and kerosene. For records indicating “others”, we consider them each individually. Based on the explanation given by the respondents, we classify these entries into either “*ph*” or “*nph*.” In our study sample, about 13.2% of the respondents use polluting heating fuel as their main or secondary heating option.

Based on the literature review and the data information, the polluting heating fuel choice may be associated with the following diseases: respiratory disease, asthma[1], allergy, and lung cancer. Our data do not provide specific information on lung cancer. Therefore in this study, asthma, allergy and other respiratory disease will be focused. About, 7.1%, 6.6% and 7.2% of the sample suffers from allergy, asthma, and other respiratory disease respectively.

---

[1] Because asthma is one of the high prevalence diseases in Kentucky (according to the CDC’s Behavioral Risk Factor Surveillance System (BRFSS) data, Kentucky ranked 2nd among the 50 states in the prevalence of adult asthma in 2002 and 15th in 2007), it is reported separately from other respiratory disease in the original data set. [22]

Counties in eastern Kentucky are located in the Appalachian Mountain range. Due to their unique geographic position, numerous past demographic, economic, and environmental studies have noticed the potential difference between this region to the rest of Kentucky. We created a dummy variable "*eastky*" to indicate whether the respondent live in the eastern of Kentucky. A total of 63% of the respondents in our sample lives in this area. Overall, 17.3% of the eastern KY respondents use polluting heating while 6.2% of the respondents in other areas do so. As a result, when we explore the impact of heating fuel choice, we need to consider the region factor.

The sample distribution, cross frequency table, and Z test results comparing the relationship between disease and pollution heating using (shown in Tables 2 and 3) provide us some direct view of the issues involved. The first column of Table 2 shows the overall distribution of polluting heating users versus non-users (13.15% versus 86.85%). The second column displays distribution of polluting heating users for those who have respiratory disease (excluding asthma). The prevalence of respiratory disease is higher within respondents using polluting heating than that of within the non-polluting heating users (8.19% versus 7.08%). However, the Z test shows that this difference is not significant. The last two columns indicate that the prevalence of asthma and allergy within polluting heating users is lower than that of within non-polluting heating users and these differences are statistically significant according to the Z test results.

Another cross frequency table for polluting heating using rate in different health condition user groups (Table 3) shows that the polluting heating using rate is lower within people having asthma and allergy than that of within those without these conditions. Moreover, Z test results indicated that these differences are significant. Z test shows that there is no significant difference between the polluting heating using rates within people suffering with respiratory disease and those who do not.

Do above results tell us individuals using polluting heating are less likely to suffer from asthma and allergy or those who have these diseases/symptoms are less likely to use polluting heating? The results in Tables 2 and 3 show the importance of considering the causality between these observations, or testing the existence of averting activity. The two-way estimation of the relationship between disease prevalence rate and the heating fuel choice should be included in the regression and simultaneity issues should be considered.

Estimation

In this study, the Logit model was used to explore the impact of the heating fuel choice together with some demographic and lifestyle characteristics on the occurrence rate of asthma, allergy and other respiratory disease. The basic equation to be estimated is

$$Y_i = f\,(age,\ white,\ male,\ eduy,\ income,\ exercise,\ smoker,\ eastky,\ ph/\ a) \qquad (6)$$

Where Y_i is a dummy variable (i=1, 2, 3), which equals to 1 if the individual suffers from one of the three diseases: respiratory disease, asthma, or allergy respectively. The explanatory variables include polluting heating using (*ph)* and some social-economic, demographic and lifestyle variables (the definitions are referring to Table 1). *a* represents averting activity and $a \geq 0$. Based on the equation, we can only capture the true exposure- response relationship when $a = 0$.

To test whether $a = 0$, we estimate a series of equations:

$$ph = f\,(age, white, male, eduy, income, exercise, smoker, eastky, Y_i) \qquad (7)$$

Where the variables *ph*, Y_i (i=1, 2, 3) and other explanatory variables are defined as above.

Results

Causality between Heating Fuel Choice and Disease

Based on the sample cross frequency table and Z test results, we did a two-way logistic regression to explore whether $a = 0$. Table 4 reports the regression results for the prevalence of the three diseases (where the dependent variables are whether the individual suffers from one of the three diseases), and Table 5 indicates the estimation results for the polluting heating choice (where the dependent variable is whether the individual uses polluting heating fuel). From Table 4, the use of polluting heating does not have significant effect on the prevalence of the respiratory disease. Using polluting heating has a significant negative effect on the prevalence of asthma and allergy. These results may be explained by the results shown in Table 5. Suffering from respiratory disease has no significant effect on the choice of using polluting heating fuel while suffering from asthma or allergy has a significant negative effect on people’s choice of polluting heating fuel.

These results may be explained by the averting activities of individuals over time. Asthma is a chronic lung disease, and allergy is the 5th leading chronic disease in the U.S (Asthma &Allergy Foundation of America, 2008). People having either of these two chronic diseases many times shift to non-polluting heating in order to relieve the symptoms of the disease. While most acute respiratory diseases – a very common branch of respiratory disease – are sudden viral infections, there is no strong motivation for people to take some averting behavior (like shifting to non-polluting heating) after the infection passes. The above results state that the causal relationship between using polluting heating and the prevalence of asthma and allergy may work in both directions and simultaneity bias is produced if just uses standard logistic regression.

There is a concern about the possibility of the simultaneity problem existing between the lifestyle variables and the occurrence of disease. In our case, the two lifestyle variables are “*excise*” and “*smoking*.” Based on literature review, we do not have sufficient evidence to support that people suffering from these three particular diseases can reduce or relieve the related health risk by changing lifestyle: exercise more or quit smoking. As a result, in this study, the possible simultaneity problems associated with the two lifestyle variables are not explicitly addressed.

GMM-IV Estimation and Results

In this study, for the case of asthma, we choose “*age, eduy and eastky*” as the IVs for the endogenous variable *ph*. All of these three IVs are highly correlated with the polluting heating using *ph* (at the 1% significance level), but do not have significant effect on the prevalence of asthma. The model of interest is:

$$\Pr(y_t = 1) = \frac{e^{X_t \beta}}{(1 + e^{X_t \beta})} \text{ or } (1 + e^{-X_t \beta})^{-1} \tag{8}$$

y_t equals to 1 if individual t suffers from asthma and 0, otherwise. X is an N × 5 matrix of regressors which include five explanatory variables ($\mathrm{K} = 5$): "*white, male, income, smoker* and *ph*." Because the model is overidentified, the GMM estimate of the vector $\hat{\beta}$ can not be obtained by solving the set of equations like equation (3). Instead, it can be obtained by minimizing the criterion function as equation (5).

Model specification for the allergy case is similar as the asthma case. "*age*" and "*smoker*" were chosen to be the IVs for *ph*. Both of these two variables are highly correlated with the polluting heating using (at the 1% significance level) and do not influence the prevalence of allergy directly (although the "*smoker*" has a marginal significant effect on the prevalence of allergy).The model of interest is:

$$\Pr(y_t = 1) = \frac{e^{X_t \beta}}{(1 + e^{X_t \beta})} \text{ or } (1 + e^{-X_t \beta})^{-1} \tag{9}$$

Where y_t equal to 1 if individual t suffers from allergy and 0, otherwise. X represents an N × 6 matrix of regressors. The explanatory variables include "*white, male, income, eduy, eastky* and *ph*." Under the overidentified context, GMM estimate of the vector $\hat{\beta}$ can be obtained by minimizing the criterion function like equation (5).

Before we estimated the model, we checked the endogeneity of the troublesome regressor. The Hausman test can be used for this purpose although it is difficult to implement for a non-linear discrete model. However, the two-way logistic regression results did provide the strong evidence to support the endogeneity of the polluting heating choice in the prevalence of asthma or allergy. We also need to test if an instrument is uncorrelated with the error term (the validity of the IV). In the overidentified case, the Sargan test can be used for this purpose. However, because it is not easy to obtain the IV residual in the highly non-linear discrete model, it is difficult to do the Sargan test for the Logit model. Based on the model specification discussed, we can obtain the GMM-IVE for the prevalence of asthma and allergy using LIMDEP 9.0 software.

Table 6 presents the estimation results of the standard logistic regression and GMM-IVE for the prevalence of asthma. In the GMM-IVE model, excluding the IVs ("*age, eduy, eastky*"), we kept all regressors in the original model except variable "*exercise*" to make the model converge better. Comparing the results of standard logistic and the GMM-IVE, we can find the coefficients estimates from both methods are identical while the standard error and the significance of the estimates (*P* value) have some differences. The standard errors of GMM-IVE are higher than the ones in the standard logistic regression. This is because less information (only a portion of the information in the endogenous variable) is used to produce the slope estimate, and the variance of the IV estimator is larger.

In term of GMM-IVE results, male are less likely to suffer from asthma at the 1% significance level which is same as the results from the standard logistic regression.

Asthma is still more prevalent within people with higher income, which may be because the prevalence record here is based on the survey question that "whether you are told by the doctor that you suffer from some certain disease." People with higher income are more likely to go to see doctor and subject to diagnosis for asthma. Smoking is not significant on the suffering from asthma, which is different from the results in the standard logistic regression. Excluded the effect of the averting behavior (that is people suffering from asthma may be more likely to choose non-polluting heating to control the symptoms), using polluting heating do not have significant effect on the prevalence of asthma, which is different from the result in the standard logistic regression.

Table 7 presents the estimation results of the standard logistic regression and GMM-IVE for the prevalence of allergy. In the GMM-IVE model, we kept all regressors included in the original model except the IVs ("*age, smoker*"). Comparing the results of standard logistic and the GMM-IVE, we can find the coefficients estimates for both model are identical except a very small difference on "*income*". The standard errors are higher in GMM-IV estimator than in Logit regression, especially on "*ph*". As of GMM-IVE results, same as in the standard regression, white American and female are more likely to subject to diagnosis for allergy at the 5% and 1% significance level respectively. Income and education have less significantly positive effect on the prevalence of allergy. Unlike in the standard Logit model, whether people live in eastern Kentucky do not have significant effect on the prevalence of allergy. Same as the asthma case, using polluting heating is not a significant determinant of the prevalence of allergy either.

Conclusions

By using the standard Logit regression, the relationship between polluting heating using and prevalence of some diseases in rural Kentucky was estimated for the period 2005-2006. The use of polluting heating fuel (including coal, wood, fuel oil, and kerosene) do not have a significant positive effect on the prevalence of respiratory disease (excluding asthma) while have a significant negative effect on the prevalence of asthma and allergy. These results may be explained by the averting behavior of individuals who shifted over time to non-polluting heating fuels such as electric and gas furnaces after they were diagnosed with asthma or allergy.

To further investigate the exposure-response relationship between the use of polluting heating fuel and the prevalence of the diseases, we conducted a reverse logistic analysis. The results show that people having asthma or allergy are less likely to use polluting heating fuel (at the 5% and 1% significance level respectively). The results suggest that people with asthma or allergy may have changed heating source over time. We used Instrumental Variable Estimation (IVE) to address this simultaneity problem and obtain the consistent estimates through the Generalized Method of Moments (GMM).

After correcting for simultaneity bias resulting from the averting behavior, using polluting heating fuel is not a significant determinant of the prevalence of asthma and allergy. There is no strong evidence to support the positive relationship between polluting heating fuel using and the prevalence of asthma, allergy and other respiratory disease.

There are some possible explanations for the above results. The lack of detailed data on historical exposure to the pollution and the use of the type of heating fuel as a proxy for the actual exposure to the pollution could be producing some measurement errors. However, a more plausible explanation related to public policy is that the performance standard promulgated by EPA and the awareness by the consumer of the possible hazards associated with different heating fuels has prompted a strong averting behavior. Information availability on energy source performance standards has allowed better informed decisions by many consumers.

Results from this study show that some demographic and personal lifestyle characteristics do have significant effects on the prevalence of the three diseases. Female are more likely to suffer from asthma and allergy. People who participate in physical activities are less likely to suffer from respiratory disease (excluding asthma) while smokers are more likely to suffer from it, holding other factors constant.

References

Amemiya, T. 1985. *Advanced Econometrics*. Harvard University Press, Cambridge, MA, U.S.A

Asthma and Allergy Foundation of America (AAFA). 2008. *Asthma Facts and Figures*. Intercepted online on Dec 10, 2008 at: http://www.aafa.org/display.cfm?id=8&sub=42

Bruce, Nigel, R .Perez-Padilla, and R. Albalak. 2000. "Indoor air pollution in developing countries: a major environmental and public health challenge." Bull. WHO 78:1078– 92

Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System (BRFSS) data. Intercepted online on Dec 14, 2008 at:
http://www.cdc.gov/asthma/brfss/02/current/current.pdf (BRFSS 2002 data)
http://www.cdc.gov/asthma/brfss/07/current/tableC1.pdf (BRFSS 2007 data)

Davidson, R. and J. G. MacKinnon.1993. *Estimation and Inference in Econometrics*. Oxford Univ. Press. New York.

Energy Information Administration (EIA). 2007. *Space Heating Characteristics by Urban/Rural Location, sub-table of Residential Energy Consumption Survey (RECS) data* (2005). Intercepted online on May 6, 2007 at:
http://www.eia.doe.gov/emeu/recs/recs2005/hc2005_tables/hc4spaceheating/pdf/tablehc8.4.pdf

Ezzati, M. and D. M. Kammen. 2001. "Quantifying the effects of exposure to indoor air pollution from biomass combustion on acute respiratory infections in developing countries" *Environmental Health Perspectives*. 2001 May; 109(5): 481–488.

---. 2002a. "Household energy, indoor air pollution, and health in developing countries: knowledge base for effective interventions." *Annual Review of Energy and the Environment*. Nov 2002, Vol. 27: 233-270

---. 2002b. "The health impacts of exposure to indoor air pollution from solid fuels in developing countries: knowledge, gaps, and data needs." *Environmental Health Perspectives* 110(11): 1057-1068.

Ezzati, M, A.D. Rodgers, A.D. Lopez, et al.2003.*Comparative quantification of health risks: global and regional burden of disease due to selected major risk factors*. Vol 2. Geneva, Switzerland: World Health Organization.

Foster, E. M. 1997. "Instrumental Variables for Logistic Regression: An Illustration." *Social Science Research* 26, 487–504

Freme, F.2009. *U.S. Coal Supply and Demand: 2008 Review*. U.S. Energy Information Administration (EIA). Intercept online on May 5, 2009 at:
http://www.eia.doe.gov/cneaf/coal/page/special/feature08.pdf

Greene, W.H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Hayashi, F. 2000. *Econometrics*. 1st ed. Princeton, NJ: Princeton University Press.

Houck, J. E., P. E. Tiegs, R. C. McCrillis, C. Keithley and J. Crouch.1998. "Air Emissions from Residential Heating: The Wood Heating Option Put into Environmental Perspective." *The Proceedings of a U.S. EPA and Air Waste Management Association Conference: Emission Inventory: Living in a Global Environment*. v.1, pp 373-384.

Kentucky Demographics. 2005. Intercepted online on Jan 2, 2009 at: http://www.eredux.com/states/demographics.php?id=1135

Smith, K.R., J. Samet, I. Romieu, and N Bruce.2000. "Indoor air pollution in developing countries and acute respiratory infections in children." *Thorax* 55: 518-532.

U.S Census Bureau. *State & County Quickfacts*.2007. Intercepted online on Jan 2, 2009 at: http://quickfacts.census.gov/qfd/states/21000.html

U.S. Environmental Protection Agency (EPA).1993. *Fact Sheet: Respiratory Health Effects of Passive Smoking*. Intercept online on May 3, 2008 at: http://www.epa.gov/smokefree/pubs/etsfs.html

U.S. Environmental Protection Agency (EPA) & U.S. Consumer Product Safety Commission. *A Guide to Indoor Air Quality—Link: What You Should Know About Combustion Appliances and Indoor Air Pollution*. Intercept online on April 2, 2007 at: http://www.cpsc.gov/cpscpub/pubs/452.html

World Health Organization (WHO).2007.*Indoor Air Pollution: National Burden of Disease Estimates*. Intercept online on July 7, 2008 at: http://www.who.int/indoorair/publications/indoor_air_national_burden_estimate_revised.pdf

Zerbe, J.I.2004. "Non-wood products/Energy from wood." Madison, WI: USDA Forest Products Laboratory, 2004. Intercept online on January 5, 2007 at: http://www.fpl.fs.fed.us/documnts/pdf2004/fpl_2004_zerbe001.pdf

Table 1. Descriptive Statistics of Variables Used in the Analysis

Variable	Mean	Median	Std Dev	Definition
age	52.828	53	14.958	continuous; age of the respondent
male	0.373	0	0.484	dummy; = 1 if male
white	0.954	1	0.209	dummy; = 1 if race is" White"
eduy	10.699	12	2.439	continuous; years of education
income	12716.550	11652	8084.590	continuous; household total yearly pre-tax income
eastky	0.630	1	0.483	dummy; = 1 if live in the eastern of Kentucky
smoker	0.526	1	0.499	dummy; = 1 if has ever used tobacco products
exercise	0.445	0	0.497	dummy; = 1 if participate in any physical activity
ele	0.669	1	0.471	dummy; = 1 if use electric as heating type
gas	0.298	0	0.457	dummy; = 1 if use gas as heating type
coal	0.034	0	0.182	dummy; = 1 if use coal as heating type
wood	0.071	0	0.256	dummy; = 1 if use wood as heating type
foil	0.006	0	0.080	dummy; = 1 if use fuel oil as heating type
kero	0.038	0	0.191	dummy; = 1 if use kerosene as heating type
oth	0.002	0	0.047	dummy; = 1 if use "other" heating type
ph	0.132	0	0.339	dummy; = 1 if use polluting heating fuel
resp	0.072	0	0.259	dummy; = 1 if suffer from respiratory disease(except asthma)
asm	0.066	0	0.249	dummy; = 1 if suffer from asthma
alg	0.071	0	0.257	dummy; = 1 if suffer from allergy
N=9539				

Table 2. Cross Frequency and Z Test Results for Rate of Disease Prevalence in Polluting/ Non Polluting Heating User Groups

Characteristic	Sample Distribution (%)	Prevalence Rate of Respiratory Disease (excluding asthma) (%)	Prevalence Rate of Asthma (%)	Prevalence Rate of Allergy (%)
Polluting Heating Users (Ph=1)	13.15	8.19	4.94	4.86
Non-Polluting Heating Users (Ph=0)	86.85	7.08	6.89	7.34
Z test Statistic		-1.426	2.620***	3.252***
P-value		0.154	0.009	0.001

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 3. Cross Frequency and Z Test Results for Rate of Polluting Heating Used by Individuals with Different Health Conditions

Characteristic	Sample Distribution (%)	Polluting Heating using rate
resp=1	7.23	14.89
resp=0	92.77	13.01
Z test Statistic		-1.426
P-value		0.154
asm=1	6.63	9.8
asm=0	93.37	13.39
Z test Statistic		2.620
P-value		0.009***
alg=1	7.02	9.12
alg=0	92.98	13.45
Z test Statistic		3.252
P-value		0.001***

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 4. Coefficient Estimates to Explain the Prevalence of Disease

Variable	Y_1= resp coeff. (Std.Err.)	Pr > ChiSq	Y_2=asm coeff. (Std.Err.)	Pr > ChiSq	Y_3= alg coeff. (Std.Err.)	Pr > ChiSq
Intercept	-4.405***	<.0001	-3.426***	<.0001	-3.929***	<.0001
	0.403		0.370		0.372	
age	0.018***	<.0001	0.003	0.292	0.000	0.981
	0.003		0.003		0.003	
white	0.366	0.151	0.269	0.225	0.522**	0.031
	0.255		0.222		0.241	
male	-0.077	0.360	-0.548***	<.0001	-0.606***	<.0001
	0.084		0.094		0.095	
eduy	-0.0318*	0.067	0.025	0.167	0.0682***	0.000
	0.017		0.018		0.018	
income	0.000	0.103	0.000011**	0.015	0.000012***	0.007
	0.000		0.000		0.000	
exercise	-0.286***	0.001	-0.152*	0.074	0.283***	0.001
	0.083		0.085		0.081	
smoker	0.953***	<.0001	0.495***	<.0001	-0.141*	0.085
	0.091		0.087		0.082	
ph	0.013	0.912	-0.313**	0.026	-0.408***	0.003
	0.115		0.140		0.139	
eastky	0.463***	<.0001	-0.109	0.208	0.186**	0.032
	0.092		0.087		0.087	
N	9539		9539		9539	
LLR	212.772		83.444		121.612	
P>ChiSq	<.0001		<.0001		<.0001	

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 5. Coefficient Estimates to Explain the Choice of Heating Fuel

	Y=ph		*Y*=ph		*Y*=ph	
Variable	coeff. (Std.Err.)	Pr > ChiSq	coeff. (Std.Err.)	Pr > ChiSq	coeff. (Std.Err.)	Pr > ChiSq
Intercept	-1.637***	<.0001	-1.619***	<.0001	-1.636***	<.0001
	0.302		0.302		0.302	
resp	0.005	0.967				
	0.115					
asm			-0.320**	0.023		
			0.140			
alg					-0.397***	0.005
					0.140	
age	-0.007***	0.001	-0.007***	0.0009	-0.007***	0.001
	0.002		0.002		0.002	
white	0.554***	0.009	0.557***	0.009	0.561***	0.008
	0.213		0.213		0.213	
male	0.073	0.255	0.064	0.324	0.061	0.344
	0.064		0.065		0.064	
eduy	-0.106***	<.0001	-0.106***	<.0001	-0.104***	<.0001
	0.013		0.013		0.013	
income	-0.00002***	<.0001	-0.00002***	<.0001	-0.00002***	<.0001
	0.000		0.000		0.000	
exercise	0.104*	0.095	0.101	0.107	0.110*	0.079
	0.062		0.062		0.062	
smoker	0.272***	<.0001	0.279***	<.0001	0.267***	<.0001
	0.064		0.064		0.064	
eastky	1.089***	<.0001	1.087***	<.0001	1.093***	<.0001
	0.080		0.080		0.080	
N	9539		9539		9539	
LLR	421.581		427.148		430.413	
P>ChiSq	<.0001		<.0001		<.0001	

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 6. Logit and GMM-IV Estimation Results for Prevalence of Asthma

	Logit			GMM-IVE		
Variable	Coeff.	Std. Err.	|P[|Z|>z]|	Coeff.	Std. Err.	|P[|Z|>z]|
CONSTANT	-3.070***	0.226	0.000	-3.070***	0.502	0.000
MALE	-0.555***	0.093	0.000	-0.555***	0.176	0.002
INCOME	0.00001**	0.000005	0.015	0.00001*	0.000006	0.061
WHITE	0.241	0.219	0.272	0.241	0.239	0.313
SMOKER	0.471***	0.085	0.000	0.471	0.891	0.597
PH	-0.356***	0.137	0.009	-0.356	1.051	0.735
N	9539			9539		

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.

Table 7. Logit and GMM-IV Estimation Results for Prevalence of Allergy

	Logit			GMM-IVE		
Variable	Coeff.	Std. Err.	|P[|Z|>z]|	Coeff.	Std. Err.	|P[|Z|>z]|
CONSTANT	-3.946***	0.314	0.000	-3.946***	0.497	0.000
WHITE	0.531**	0.241	0.028	0.531**	0.249	0.033
MALE	-0.645***	0.093	0.000	-0.645***	0.098	0.000
INCOME	0.000013***	0.000004	0.004	0.000011*	0.000006	0.065
EDUY	0.073***	0.017	0.000	0.073**	0.029	0.013
EASTKY	0.211**	0.086	0.014	0.211	0.232	0.363
PH	-0.397***	0.138	0.004	-0.397	2.952	0.893
N	9539			9539		

*, **, and *** represent significant at the 10%, 5%, and 1% significance levels respectively.